



Alaska Alternate Assessment New Mentor Rater Reliability

Revised

May 16, 2011

Inter-rater Reliability
Training to Standards
New Mentor Training

On September 27 and 28, 2010, Alaska educators seeking certification as qualified assessors (QA) gathered in Juneau, Alaska. These assessors-in-training (AIT) participated in a half-day of on-line orientation, and a half-day of on-line web-based training, followed by web-based proficiency testing. AITs were expected to complete all training and proficiency testing prior to joining the All Mentor Training in Anchorage at the end of October.

The second day of New Mentor Training (September 28) focused on supervised administration of practice tests to adults. There were ten new assessors-in-training and six returning qualified assessors. Assessors-in-Training who are new to the Alaska Alternate Assessment are required to administer practice tests, in order to earn Qualified Assessor status.

New Mentor Training, Practice Tests

New assessors administered practice assessments to participants who had previously acquired Qualified Assessor status. In some cases, new assessors administered the practice tests to other new assessors. Each of the new AITs administered one practice test in Reading, Writing, Mathematics and Science, as well as items from the ELOS Math and ELOS Reading administration manual. Each trainee was also evaluated on the overall administration of the practice assessments.

In each subject area, percent correct were calculated for appropriately recording the student's response (Student Response Total and Student Response Percent) and for correctly recording the appropriate score earned by the student (Score Record Total and Score Record Percent). The results are displayed for each subject area.

In each subject area, a new Mentor is asked to correctly record the student response in the scoring protocol. If the Mentor correctly records the student response, a score of 1 is recorded in the calculations. If the Mentor does not correctly record the student response, a score of 0 is recorded. For each subject area, a total possible Student Response is tallied (1 point for each correctly recorded response) for all new Mentor practice tests. The total is recorded in the included tables, as "Student Response Total."

The same tally is derived for total Score Record, an assessment of whether the new Mentor correctly scored the item in the scoring protocol. For each subject area, a total possible Score Record is tallied (1 point for each correctly scored item) for all new Mentor practice tests. The total is recorded in the included tables, as "Score Record Total."

Student Response Percent is a calculation of the total number of correctly recorded student responses per subject area, divided by the total possible Student Responses. Score Record Percent is a calculation of the total number of correctly scored responses, divided by the total possible scores.

Total Possible and Total Percent are calculations of Student Responses and Score Records combined, for a total percent inter-rater agreement per each subject area.

Practice Test Results

Total agreement among assessors in Reading was 0.98.

Reading	Practice Test
N	16
Possible points	9
Student Response Total	138
Score Record Total	144
Student Response Percent	0.96
Score Record Percent	1.00
Total Possible	288
Total percent	0.98

Four Assessors in Training were required to rescore and resubmit their protégé analysis of the Writing assessment. Total agreement among assessors in Writing was 0.93.

Writing	Practice Test
N	16
Possible points	7
Student Response Total	106
Score Record Total	103
Student Response Percent	0.95
Student Response Percent	0.92
Total Possible	224
Total percent	0.93

Total agreement among assessors in Mathematics was 0.97.

Mathematics	Practice Test
N	16
Possible points	14
Student Response Total	216
Score Record Total	218
Student Response Percent	0.96
Score Record Percent	0.97
Total Possible	448
Total percent	0.97

Total agreement among assessors in Science was 0.91.

Science	Practice Test
N	16
Possible points	8
Student Response Total	116
Score Record Total	116
Student Response Percent	0.91
Score Record Percent	0.91
Total Possible	256
Total percent	0.91

Some sections of the ELOS practice tests were missing from the packets of three Mentors-in-Training; their scores are not included in the percent calculations for this section. Total agreement among the remaining thirteen assessors in ELOS administration was 1.00.

ELOS	Practice Test
N	13
Possible points	5
Student Response Total	65
Score Record Total	65
Student Response Percent	1.00
Score Record Percent	1.00
Total Possible	130
Total percent	1.00

Overall Review scores are an assessment of an AIT's or QA's adherence to test administration and scoring protocols, with possible scores of 1 = Unacceptable; 2 = Needs Additional Work; 3 = Satisfactory; 4 = Exceptional. AITs and QAs averaged a total percent agreement of 0.95.

Overall	Practice Test
N	16
Possible points	28
Total Earned	426
Total Possible	448
Percent	0.95

New Mentor Training, Review of Protégé's Test Administration

All participants in the training (AITs and returning QAs) participated in a review of a protégé's test administration in each of the four subject areas and in ELOS administration. Inter-rater reliability scores were determined based on an analysis of participants' reviews of a standardized protégé assessment administration packet of tests in Reading, Writing, Mathematics and Science, and ELOS administration. Additionally, participants were scored on overall administration constructs of accuracy in cover page, task administration, task scoring, clarity of markings and appropriate choices to stop testing.

In each subject area, percent correct were calculated for appropriately recording the student's response (Student Response Total and Student Response Percent) and for correctly recording the appropriate score earned by the student (Score Record Total and Score Record Percent). The results are displayed for each subject area.

Review of Protégé's Test Administration Results

Total agreement among assessors in Reading was 0.78.

Reading	Protégé Review
N	15
Possible points	9
Student Response Total	79
Score Record Total	132
Student Response Percent	0.59
Score Record Percent	0.98
Total Possible	270
Total percent	0.78

Four Assessors in Training were required to rescore and resubmit their protégé analysis of the Writing assessment. Total agreement among assessors in Writing was 0.73.

Writing	Protégé Review
N	15
Possible points	7
Student Response Total	97
Score Record Total	56
Student Response Percent	0.92
Score Record Percent	0.53
Total Possible	210
Total percent	0.73

Total agreement among assessors in Mathematics was 0.82

Mathematics	Protégé Review
N	15
Possible points	14
Student Response Total	164
Score Record Total	179
Student Response Percent	0.78
Score Record Percent	0.85
Total Possible	420
Total percent	0.82

Total agreement among assessors in Science was 0.85.

Science	Protégé Review
N	15
Possible points	8
Student Response Total	95
Score Record Total	109
Student Response Percent	0.79
Score Record Percent	0.91
Total Possible	240
Total percent	0.85

Total agreement among assessors in ELOS administration was 0.77.

ELOS	Protégé Review
N	15
Possible points	5
Student Response Total	53
Score Record Total	62
Student Response Percent	0.71
Score Record Percent	0.83
Total Possible	150
Total percent	0.77

Total agreement among assessors in Overall administration was 0.71.

Overall	Protégé Review
N	15
Possible points	28
Total Possible	420
Total	298
Percent	0.71

Overall Review scores are an assessment of an AIT or QAs adherence to test administration and scoring protocols, with possible scores of 1 = Unacceptable; 2 = Needs Additional Work; 3 = Satisfactory; 4 = Exceptional. AITs and QAs averaged a total percent agreement of 0.82.

Overall	Protégé Review
N	15
Possible points	14
Student Response Total	164
Score Record Total	179
Student Response Percent	0.78
Score Record Percent	0.85
Total Possible	420
Total percent	0.82

Implications for Future Training

The writing task 1.56B: Write Own Name presented the greatest challenge to Assessors-in-Training, with only eleven of sixteen assessors recording the score correctly in the practice test. Greater time with enhanced examples should be scheduled for future training. Scoring and recording for Correct Letter Sequences, and Correct Word Sequences also presented a challenge. This area has been a focus in training, and will continue for future training.

Additionally, the task of reviewing a protégé's work was problematic, with no section earning greater than 82% agreement among trainees. This system should be explicitly demonstrated and practiced during the training. Perhaps the trainers could schedule each subject's practice test and protégé evaluation as a group. The trainees would administer the practice tests and then could review their own work carefully, guided by the trainers, as though their work was a protégé's. Then the protégé evaluation tools and process could be practiced. Grouping the two tasks by subject area would allow the trainers to address questions after the practice test before allowing the trainees to evaluate a protégé's work.